Agency in socio-technical systems

Problem statement

Developments in the field of artificial intelligent suggest an increase in the agency of machines, as we are assigning them roles and control that would have been in human hands. However, the unpredictability of the actions undertaken by artificial agents while they are not fully controllable by a human leads to situations where agency becomes an issue (1). Scholars have long debated the impact of machine agency on the diffusion of responsibility. While attributing responsibility is the social function of being in control (i.e. sense of agency) (2), current artificial agents cannot be held responsible.

While much of the technology for such agent is at hand, the social-cultural aspects of the human robot interaction have yet to be addressed further. After all, the objective capabilities of these machines are less relevant than the subjective willingness of human agents to accept this collaboration (3). All these arguments imply a fresh look at concepts of machine agency. Currently, there is only anecdotal evidence about the impact of robots in the society and the concept of machine agency has not been clearly defined (4)(5).

Research Question

This dissertation aims to understand the effects and differences of agency ascribed to machines and analyze factors that facilitate or hinder the acceptance of artificial agents by investigating the following research question:

How is (social) agency ascribed to robots?

How can artificial agents be integrated effectively into social systems?

Ongoing Results

Autonomous robots are agents in a manner that they do things rather than have things happened to them. They have the ability to “think/plan/act” to the extent without external control. By integrating these reviews, we suggest the following definition:

“machine agency is the capacity attributed to machines to evoke changes within a socio-technical system by autonomously carrying out a goal-oriented task.”

Based on the proposed definition of machine agency, we can conceptualize artificial agents in four different types. Table I summarizes the main differences among them. Though these agents are technology-based systems, major differences exist among them in terms of how they control the input-output cycle, pursue the goal and the extent to which they can perform tasks independently.

Given that how a robot interacts with people can affect the efficiency of collaboration (6), we conducted a user study in which a social robot assists people while building a house of cards. Data from the experimental study revealed that People interacting with a robot with a person-oriented interaction style reported a higher self-efficacy in HRI, perceived higher Agreeableness of the robot and found the interaction less frustrating, as compared to a robot with a task-oriented interaction style. This suggests that a robot’s interaction style can be considered as a key factor for increasing people’s perceived self-efficacy in HRI, which, we argue, is essential for establishing trust and enabling collaboration between humans and robots.

Next is to investigate how a strengthening of machine agency, for example through increasing levels of autonomy, affect attitudes towards robot.

References

1) Weber, R., Pereira, Â. G., Dechesne, F., Timmermans, J., Van, R. K., & Lehn, H. Vom. (2013). Fact sheet - Ethics Subgroup IoT - Version 4 . 0. Delft University of Technology Chair Ethics Subgroup IoT Expert Group, 1–21.
2) Frith, C. D. (2014). Action, agency and responsibility. Neuropsychologia, 55(1), 137–142.
3) Echterhoff, G., Bohner, G., & Siebler, F. (2006). “Social Robotics” und Mensch-Maschine-Interaktion. Zeitschrift Für Sozialpsychologie, 37(4), 219–231
4) Jia, H., Wu, M., Jung, E., Shapiro, A., & Sundar, S. S. (2012). Balancing human agency and object agency: an end-user interview study of the internet of things. Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp ’12, 1185–1188
5) Lin, P., Abney, K. and Bekey, G. A. (2012) Robot ethics.
6) Schulz, Kratzer & Toussaint (2018). Preferred interaction styles for human-robot collaboration vary over tasks with different action types,” Frontiers in neurorobotics, 12,36

Setareh Zafari
TU Wien | Institute of Management Science
Department of Labor Science and Organization
setareh.zafari@tuwien.ac.at

TRUST ROBOTS